DOCUMENT RESUME

ED 226 061 TM 830 104

AUTHOR Kelley, E. W.
TITLE Standardized Tests and Public Policy: The Politics of Selection in America.
PUB DATE Sep 82
NOTE 30p.; Paper presented at the Annual Meeting of the American Political Science Association (Denver, CO, September 2-5, 1982).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070) -- Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Entrance Examinations; Court Litigation; *Educational Opportunities; Elementary Secondary Education; Policy Formation; *Politics; Professional Personnel; *Public Policy; Selection; *Standardized Tests; *Test Use

ABSTRACT

The way test scores are used by both the public and private sectors, and the way in which this use is promoted and regulated indicates how some public policies are formed and who characteristically bears the cost of policies as well as who benefits. What has been happening in the relationship between test use and policy formation is not unique, particularly when professional groups or guilds are involved. The displacement of costs, the disaggregated use of tests, the delegation of public authority, particularly at the state level, and the legislative deference given to professional judgments are inconsistent with the essentially political character of deciding who is to get what through the public sector. The use of standardized tests to implicitly define criteria of selection for educational opportunity or jobs essentially delegates a political or ethical decision to a group of individuals who, while not meaning to make that decision, project their own perception of professional norms. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

ED226061

STANDARDIZED TESTS AND PUBLIC POLICY:

The Politics of Selection in America

E.W. Kelley
Department of Government
Cornell University

Prepared for delivery at the 1982 Annual Meeting of the American Political Science Association; Denver, Colorado; September 2-5, 1982.

Partial support for the preparation of this paper was provided by the Jonathan Meigs Fund.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E. W. Kelley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 104

Western societies measure themselves both in the aggregate, as when we compile gross or net economic statistics, and comparatively, as when we measure the educational or income attainments of some individuals or segments of our population against others. Very often the results of these measures or tests are used in public policy debates. Low rates of growth or high inflation are thought to be bad. The argument is that the government should do something to ameliorate the situation. The argument is not all that clear, however, as has been known since the Constitutional Convention gave control over the circulation of currency and bankruptcy to the federal government; inflation does not hurt everyone. In a similar vein, it is not clear that declining SAT scores or static results from nationally given achievement tests hurt everyone, even though these figures are sometimes used to promote or justify public policies in the area of education. In particular, for example, the minimum competency exams mandated by at least thirty-three state legislatures or executives (1978) (Olney, pp. 1-3) are claimed to be a response to declining educational standards as signified by these declining or static test scores.

In the area of public policy evaluation, standardized and other testing is often used to advance either the argument that a policy has been good or that it has not been good, and hence, ought to be terminated or changed. However, the information concerning testing for evaluative results does not really stand on its own and is not necessarily a factor in policy continuation or change. Consider the reauthorization of Title I in 1974 and 1975. At that point, very few evaluations of Title I programs which used achievement of other standardized testing effects could produce

a positive result (early draft: Evaluating Compensatory Education: Interim Report). There are at least two reasons for this. The first is that Title I was not necessarily geared in any particular school district to raising test scores. Ironically, if test scores had increased significantly, students would cease to be Title I eligible. In addition, Title I was really about 13,000 separate programs in each of the school districts in the United States. When studies were done with aggregate data, the testing results were often misused. The purpose of Title I insofar as it related to test scores at all, was not to raise the aggregate achievement test score in a school district. It's purpose was, at least conceptually, to help a body of underachieving children achieve above a particular threshold. If one wanted to use test scores as an evaluation of the program, the extent to which this threshold was passed would have been appropriate. Of course, neither Congress nor local school districts ever defined the threshold of success, except as some level had to be set for purposes of passing out funding, given the money available. (Incidentally, where intervention programs have attempted to teach people to perform better on achievement or even aptitude tests directly, there have sometimes been reports of success. Lazar et. al., 1977). One could question whether a test of academic achievement of the normal variety is an appropriate evaluative tool for Title I; if so, which test and how evaluated? There is a question whether any one or small set of tests could be consistently applied to evaluate what were a very diverse set of programs?

The lack of any demonstrable educational results, and by this we mean achievement test scores or increases, did not prevent the reauthor-

ization of Title I. That is because the testing as measured by these tests was not the criteria the Congress used to initiate the program or to vote for its continuation. It was clear in talking with both Congresspeople and staff on the House Education and Labor Committee that a financial distribution funded on a reasonable stable basis could not be easily withdrawn. The questions were largely ones of fine tuning: should different bases of allocation be used? To withdraw the money would have been difficult then (as President Reagan is finding it difficult now in the particular case of Title-Chapter I) principally because the monies formed a dependable part of the school districts' budget. At the local and Congressional level, educators and property tax payers are sufficiently well organized that withdrawal of external money would have clearly negative electoral effects on whomever might vote for it. Indeed, in New York State, in supplemental appropriations before every biennial election, additional funds for local school districts is a centerpiece. In the case of Title I then, the evaluations using standardized tests did not speak for themselves; in fact, they were on the whole irrelevant to the causes that underlay the initiation of the policy or the programs' perpetuation. This is reflected in the kinds of information given the House Committee in the Title I study's final report (NIE, 1978; Educational Amendments of 1974; Congressional Record, 1966, pp. 25027-25581).

If standardized test scores, delining SATs, etc. do not speak for themselves in the formulation or evaluation of public policies, and if they are sometimes irrelevant for the initiation or continuity of policies, what can be said about them? They are tools like any other tools used to make more microlevel decisions. They are also tools used to justify or attack programs at all levels of government. What we can do

is to turn the question slightly on its side and ask what the use of standardized tests both by private parties and by public authorities reflects about the formation of public policies in the United States.

Public policies are decisions made by the public sector (at some level) to spend money or allocate opportunity, whether through a transfer or service intermediary, to individuals or groups in the polity. The allocation of opportunities to some often prevents others from obtaining similar opportunities, particularly in cases where individual educational or other credentials are prerequisites for a job. One feature of most public policies is that some or a lot of people bear costs and some or a lot of, but not usually the same, people obtain perceived benefits. Very seldom is any policy perceived by everyone to be a benefit. Hence, arguments using standardized test results, like arguments using "the public interest" do not carry much weight with some individuals when we ask them to behave in a manner inconsistent with their self-interest. Looking at the way test scores are used by both the public and private sectors and the way in which this use is promoted and regulated or not, can tell us something about how some public policies are formed and who characteristically bears the cost of policies as well as who benefits.

We will suggest that public policy formation involving professional groups as service intemediaries is initially disaggregrated and controlled by private and public sector organized users. Costs are displaced onto the least or most recently organized groups or individuals. Costs may be denial of choice or opportunity as well as being on the wrong end of a system in fiscal non-equivalence. The latter occurs when the public sector is involved in the passage of legislation: which legislation

then delegates authority to the professional group to regulate its existence. Finally, all developed or developing professional groups or "guilds" expect and give deference to the "professional" judgements of others, even where additional, ancillary public policy formation is involved.

Take as an example the development of minimum competency tests. At least thirty-six states now have some form of test in use or in preparation (Ewing, 1979). The advocacy and use of such tests appears to be based upon the desire to hold educators responsible for their products, to measure the cost-effectiveness of education per se, and to ensure that all students obtain an adequate eduation in basic skills (Gonnan and Rubin, 1977). The educational and testing activity associated with proficiency standards is usually seen as a cost-benefit sort of undertaking. In part it is, but to concentrate on total costs and benefits will mask the political process that is involved in standard setting, and will lead to mistaken predictions and recommendations as to the eventual character of such standards. The politics of standard setting is not determined by totals of costs and benefits. What is important to know is: whose costs and benefits? Legislatures do not pass laws and administrators do not design programs in a vacuum.

Standard setting itself is not simply a program developed by educational professionals to improve the education of children and accountability in general. Rather, it is a response to particular political and fiscal stimuli. High real property and personal income tax levels for some groups and the problems some middle-class kids have getting into professional schools lie behind the cry for accountability.

Despite some remaining slippage, access to employment and income

possibilities depend more and more on the meritocratic system. Such a system measures people, in part, by years of schooling completed and certificates in hand. The information costs of such a screening system to employers is low. Opportunity to pass through the system depends in part on the use of test scores; hence, at the level of the individual the consequences of private decisions concerning the use of test scores can still be rather substantial. Entrance to a college, professional or law school of one's choice depends in large part on how well one takes standardized tests. Insofar as there is slippage in this system, the beneficiaries are those members of minority groups who do not test as well but who have placed their case successfully on the political agenda. The net effect is that the children of the well off can have little hope of directly inheriting their parents standing in society. It is not just the parents of children who cannot read who are challenging the advocacy of American education. Critical to pressure placed on legislatures are the parents of children who cannot read or reason well enough to test into high status graduate or professional schools.

For many, the Scholastic Aptitude Test score decline of 49 points is seen as a warrant or a claim that schools are generally failing to teach kids. However, the use of this evidence to claim that basic skills are not adequately taught and to introduce minimum competency standards is wrong. SATs do not measure basic skills. If you can even read the directions for taking the machine scored SATs, you are literate enough to perform almost all the non-professional roles in our economic system (see Farr and Olchavsky, 1980).

Why don't middle-class non-beneficiaries respond similarly to the clearly inefficient system of local-state-national welfare, public

housing and rent controls, or the woeful redistributions to professionals in Medicare and Medicaid? The latter is less objectionable because it primarily redistributes from the middle-classes back to the middle-classes; middle-class tax dollars are paid for services provided by middle-class professionals. The guild-like prerogatives of doctors and hospital administrators are not threatened. Taxing is separated from spending within the large bureaucratic labyrinth of the Federal Government. Further, anyone might need Medicare when older. Public housing funnels money to the construction industry; rent control is temporary and/or subject to modification under political pressure. Further, its beneficiaries are themselves often organized - e.g. labor union members. In the cases of medicine and housing, then, there is not compelling pressure only on the side of undoing the programs. Organized middle-class groups are also beneficiaries.

Welfare is a predominantly national program involving Federal taxation and state expenditures. The perceived connection between welfare and income tax levels is somewhat tenuous because so many other activities are also funded out of this revenue. However, education is viewed as a local good, even if some of the money comes from the taxing and spending mechanisms of state and Federal governments. Middle-class individuals readily see the process of local education in the schools they walk by on their way to work. Further, half of all of what a small area or district spends on education is raised within exactly the same area through local property taxes. There is a directly visible tie between taxing and spending in a small area when the tax district is the same as the benefit district.

Politicians have long since known that when additional spending

is desirable, one must avoid the geographical identity of small taxing and spending districts. The greater the equivalence of these districts, the less is spent. The smaller the area of equivalence or overlap, the less is spent. Indeed, we must use moral suasion at the local level to spend at the level we do for education. Such suasion is not necessary for federally-funded welfare even though welfare would seem to be no more of a morality engaging issue than is education. Taxing for welfare is largely federal; most spending is state and local. Little financial control rests in a local area alone.

The activity of the middle-class tax payers begins with what they can see and feel they can control: education. This reaction takes the form of questions and doubts, initially vaguely posed: what are we getting for our money? Shouldn't we have property tax relief? etc.. These questions must be taken seriously because those who ask them are or can be more politically efficacious than many others in our society. In other forums, as doctors, lawyers, business people, etc., they are already efficacious. (See, for example, Gallop, 1977; Brodinsky, 1977; Mattleman, 1979; Hildebrand, 1980; Fisher, 1978; Gawelti, 1978).

The movement from a concern for the outcomes of formal education and middle-income tax revolt to something like proficiency standards in public education is conditioned by two factors - the cost accounting basis of much of American business life and the fact that the <u>response is conditioned by professionals in the area of education</u>. We do not see the wholesale dismissal of teachers or hear cries that too many people teach remedial reading and too poorly; instead we become concerned with the costs and benefits of education. Let's see what we get for our dollar, particularly since we are not going to spend dollars at a higher

rate. Let's try to measure the product. Of course, the people who will do all this are the professional researchers and administrators of the educational system.

States use minimum competancy exams for a variety of purposes. Some attach diplomas to successful passage of the exam. Some give certificates for passage of the exam and high school diplomas for school attendance. In California, use of the exam for graduation is a local option. There are a variety of patterns states follow. In almost every case, though, where the state legislature has acted, it has adopted a "general intent" bill and left the determination of what constitutes an adequate education or competancy to someone else. In the case of Florida, the initial minimum competancy exam was hastily prepared after the state legislature simply mandated some kind of such exam. In Debra P., the federal court outlawed use of the test not because the test was hastily prepared, not known to measure competancy, and not known to be predictive of any of the activities for which education prepare one, but because it was being applied before the last students who started in Florida's formerly dual system of education had passed through high school. The courts have taken the use of test themselves and the adequacy of tests to measure what they claim to measure more and more to be matters of professional judgement to which they defer. No state legislature has yet detailed for purposes of the establishment of an exam, exactly what it should measure, let alone insist that whatever exam or threshold is used actually does adequately measure or predict the measure of these tasks.

Clearly, educators are not disadvantaged by this legislation. It is impossible to hold a school district accountable for the successful

passage or the percent of its students that pass this exam unless the assumptions of no mobility and constant ability are met. Additionally, individual teachers cannot be held accountable very easily. If teacher accountability were desired and were possible in today's unionized environment, the once or twice yearly achievement tests already given would be adequate measures of growth for purposes of measuring what teachers teach. Certainly they are sensitive over a broader range than are minimum competancy exams, which are designed to be sensitive particularly around the point which distinguishes passing and failing. Politicians are doing something about education, but that something is not redistributive. To respond to the difficulties that middle-class parents have when their children do not obtain adequate educational placement would be self-defeating. Every child placed in a competitive program implies that there is one child who must be taken out. However, everbody can be competant. Yet, students who work hard and persist in a classroom but don't test well, are moved further down the education and job queue. There is cost to the system, but it is borne by those least organized and typically, most disadvantaged.

An obvious consequence of the preceeding narrative is that the cost of this public policy involving educational competancy is most clearly borne by those least able to afford it and/or most recently organized or unorganized. These are the disparate students who would work but yet fail to pass the exam; they would have no clear organizing focus. Almost no organized group associated with the development of minimum competancy tests is a net loser. A similar process occurs when one uses standardized testing for selection for educational opportunity or jobs. A clear set of beneficiaries are those people doing the selecting. So

long as the testing devises themselves are non-perverse, (by that we mean they rank order candidates at least randomly with respect to performance on what is being selected for), those who pay the cost are those who pay for the test, the students. The students also are those who benefit or suffer the consequences of the use of the test. Additionally, developers of standardized tests as well as various professional groups in higher education are net beneficiaries of the use of these tests.

In this instance, however, former test takers and middle-class parents have successfully passed "truth-in-testing" legislation. This legislation does not at all control the uses of tests nor does it control the content of tests. All it does is divulge information to the test-takers and general public concerning the purposes and content of the tests. Even the latter was formally held a secret and test producers lobbyed very hard to attempt to maintain that information. The production of information, however, by itself has not changed the costs and benefits of the use of the tests to various groups and individuals; information is necessary for those who bear costs to affect change, however.

The capacity of testing professionals, test producers, and users to impose costs on those who take tests is clearly evidenced in the institutional responses to New York's legislation. Most, but not all universities in the state were weakly opposed to the "truth-in-testing" bill. Letters were sent to the Joint Committee on Higher Education, in which they deferred to the opinions of testing "experts" employed by test producers in an almost <u>verbatim</u> manner. Little follow-up occurred: the legislation did not restrict the options of users. After the legislation passed, some test producers and/or intermediaries withdrew tests temporarily from N.Y. State. Schools still used tests; students

needed to take them. Unilaterally, test producer increased costs to students in N.Y.S.. N.Y.S. schools could still fill their classrooms. N.Y.S. students had to go out of state in some cases to qualify for these classrooms. Test producers placed blame on the legislation; tests could not be properly administered and scored under the new law. Both the weight of the testimony on the bill, then, and current practice (1982) show that this professional claim concerning the effects of the legislation was incorrect. (see: Testimony of Allan Nairn, 1979; letter of Cornell U., 1979; letter of Law School Admission Council, 1980; testimony of Stephen D. Solomon, 1979; Memorandum of Opposition, ETS, 1979; Memorandum in Opposition, Associated Medical Schools, 1978; Memorandum of CEEB, 1979.).

When standardized tests are used by a variety of public and private sector agencies, the low cost of this use and the displacement of the cost to those individuals providing the data provide another mechanism of fiscal non-equivalence. Again, the more organized groups benefit. Costs are displaced on the least organized even after legislatures act, as in the New York State case. The absence of policy allows for disaggregated, but organized, interests (testing companies, professional schools, universities, etc.) to benefit from a sequence of small fiscal inequivalencies. Displacement of cost, then, and disaggregation of policy sources is the starting point for American public policy when existing or nascent professional groups are involved.

Grant McConnell and others have well documented the delegation of the public authority to private groups through legislation and administrtive cooperation or cooptation. In the case of federal policy towards agriculture and labor there developed an extensive federal bureaucracy

which shared a mutual or symbiotic interest with organized private groups in ensuring that public sector benefits flowed to those groups. Within the public sector there was created a natural monopoly for purposes of distributing the good or service to these groups; the groups, whether they be farm bureaus or labor unions, themselves represented monopolies with regard to membership and the distribution of benefits from the public sector. Temporally between the development of business firms within which free-riding is difficult, if not impossible, and the development of a public sector with a symbiotic interest in providing private sector internalities, organized groups of professionals developed. These groups, akin in many ways to medieval guilds and, in Britain and Switzerland, their successors, also attempt to attain delegation of public authority for their private purposes. In the United States, professional groups typically prefer to be unregulated unless competition of a substitutable form comes onto the scene. The professional group then typically goes to <u>state</u> legislators to obtain both a monopoly status and a delegation of public regulatory authority. The history of the AMA and the various state legislatures in resisting the onslought of both the osteopaths and the homeopaths is well known (Burrow, 1977 Part 2). The evolution of state bar associations in the early part of this century (after legislatures had emaciated the colonial associations between the revolutionary and civil war) which, when combined with admission practices of the law school, restricted entry into and terms of practice of the legal profession has also been documented. These groups and others (like accountants) trade a delegation of public authority for a guarantee of quality of service. It is expected that in the areas of their claimed expertise, deference will be given to their judgments. As a

general rule, this public sector delegation of autnority does not always allow a professional association to fix levels of remuneration and/or the availability of the service, both quantity and distribution. However, there is a clear tendency this way which has only recently been challenged as the consumers of services provided by professionals have themselves become organized.

One can conceive of professional educators associated with testing and evaluation, and the producers of such tests and evaluative instruments as being collateral nascent guilds. In almost every case in which minimum competancy exams have been legislated, the legislature has deferred all particulars of the content of such tests and implicitly, what it means to be competant, to the judgment of professional educators both within and without the public sector. In testimony in New York State during the process of enactment of the "truth-in-testing" legislation, officers of at least one producer of standardized tests claimed that deference to expert judgment as to what constituted a good and bad test was appropriate. So long as the test results are nonperverse, both public and private sector users will gladly defer to these claims. In other countries, particularly Brazil and Mexico, the public sector as well as the private sector has taken explicit cognizance of such professional claims and given their approval. Educational advancement to the university level in those countries and in a number of others is even more rigidly governed by the systematic application of test scores in the admissions process than it is in the United States (Harrell, 1970; Havighurst and Gouveia, 1969; Kneller, 1951). In the United States, the extent to which test scores are used for puposes of educational admissions and selection for employment is up to the individual user. Remember,

this disaggregation of judgement and effective policy determination is a feature of American public policy generally.

Recognizing that it is impossible to get a legislature to easily restrict what are the historic prerogatives, particularly of the private sector, in the use of tests, the advocates of "truth-in-testing" legislation advocated only that information be made available to students and accuracy of scoring be guaranteed. One could not take on all the professions that conveniently used test results in professional school admission processes as well as most of that part of the private sector that uses the scores for purposes of selecting employees. Costs to these groups could only have increased if legislation attempted to define equity or specify restrictions in the use of tests. The groups advocating "truth-in-testing" do not have sufficient floating votes or other resources appropriately located to enter the system of Congressional reciprocity the way traditional minorities have. (Note: this also explains why slippage in the use of tests for selection benefited these minorities.).

The only area in which the production and use of standardized tests is internally and professionally regulated is in the area of bias. The term "bias" has become rather equivocal in use. Origianlly, it meant simply that individuals were chosen for positions of education opportunity for reasons that were not consistent with legislative public policy or were not consistent with certain norms or ethics that we publically claim to share. The testing profession has altered the use of that term to mean a number of different things that are characteristics of tests themselves.

The first point to understand about bias is that it is defined in

terms of comparisons. A test is never simply biased. It is biased only when the scores on tests or individual questions for one group are compared with those for another group. Since test-takers can be divided into groups many ways (tall, short; Black, Hispanic, White; male, female; etc), tests may have bias in some comparisons and not in others. The comparisons we actually make are usually determined politically or otherwise outside the domain of test construction or evaluation itself.

Bias comes in as many forms as there are writers on the subject, yet we can probably breakdown uses within the testing profession into four categories (Maw, 1979). Some literature on test bias simply asserts that tests (and sometimes test items) are biases because some group of examinees uses the language in which the questions are posed more frequently than other groups (Barnes, 1969; Hall and Turner, 1974; Hill, 1977; Williams, 1971). A variation on this position is that differences between groups can occur because of different familiarities with format and test procedure (Flynn and Anderson, 1976; Thoman, Hertig, Dryman, and Fernandez, 1971; Evans and Reilly, 1972). The latter sources of bias are presumably accounted for by variation in procedure or training groups of examinees in test-taking formats. The former, linguistic, source of reputed bias is more difficult to handle. Item content can be changed to include words equally familiar to members of different groups of examinees. Sometimes, and mistakenly, it is assumed that some skills are best measured in language-free tests (i.e., Raven's Progressive Matrices). Exactly the opposite is true; the vocuabulary required in the thinking process which traces the rotation of figures, lines of symmetry, etc. is highly culturally loaded, found often in the habitually used language of some groups and rarely in others (Cohen, 1969).

In any event, such tests are hardly relevant when one wishes to test a student's practical reading skills.

A second measure of bias can be labelled predictive (Potthoff, 1966; Linn, 1973; Peterson and Novick, 1976; Rudner, 1977). The predictive validity of a test across groups must be the same if a test lacks bias. Otherwise, the scores of individuals within some groups are better predicted than those in others, and it is unfair to use the test results as if they are equally accurate across all groups. The methods of prediction vary and include standard regression and conditional probability models, but the logic is the same in all cases.

The third category of test bias is based on models of consistency. While each item is not assumed to be equally difficult for all groups, any item is to be as relatively difficult compared to another item across all groups for a test to be unbiased. Otherwise, an item, for whatever reason, is relatively more difficult for at least one group, even considering the possibly different levels of skill in the tasks being tested in the different groups (Cardell and Coffman, 1964; Cleary and Hilton, 1968; Lord, 1977; Merz, 1973). The models employed in analyzing this type of bias also vary, but include both ANOVA and factor analysis as well as item characteristic curve analysis. In all cases, regardless of criterion, lack of consistency is equated with bias.

Even if all the above sources of bias have been controlled, there remains another possibility: distractor bias in a multiple choice format (Maw, 1979). Members of some groups may be particularly attracted to some distractor items on multiple choice tests. Spanish speaking students, for example, may be particularly attracted to capitalized noun distractors. The possible causes of such bias are probably almost

limitless. Again, like the previous consistency models, one assumes groups to be attracted to different distractors at proportionately the same rate. Where this is not the case for one group (compared to the others), distractor bias exists and lowers the frequency of correct responses in such a group. Distractor bias is particularly pronounced in timed, multiple choice tests, and is particularly unfortunate when totally unrelated to what the test is about.

Clearly, considerations of equity require that some kinds of bias in such tests be reduced if not eliminated. Format and procedure bias can be checked directly. Linguistic bias is likely to appear in one of the other three categories given above. Predictive bias can be caused in part by bias due to lack of item consistency or distractor bias, although both of the latter may remain even when there is no predictive bias. Bias defined in terms of item consistency or distractors has the advantage that, when found, the items or distractors that are the cause of bias are identified and can be changed.

The danger of considering only these forms of bias is clear. We have no particular reason to believe that standardized tests (particularly of "aptitude") that meet all these conditions necessarily select the best people for an educational opportunity or for a job. The tasks of everyday life and job performance are not brought into the test. Instead, the test obtains some face validity by asking about the task.

Whether a test predicts performances of value is a political issue, not an issue within testing itself. Further, a test may predict performance but it might not be performance of the kind that an individual is being selected for. Very often as a consequence of the use of tests, we select people for training for a job, not for the jobs themselves.

Then, the adequacy of the test becomes somewhat self-fulfilling. Clearly, those people who are trained say, to be a department store buyer or an accountant, are better at those tasks than those who aren't.

Remember, though, standardized tests were never designed to maximize individual equity. The great benefit of a test and a test score is that it provides some rank-order which can be used for low-cost selection. The cost of selection using information like test scores compared across individuals is very low compared to the cost if one were to use less quantitative materials. From the point of view of the educator or employer, tests are highly efficient.

One might argue that if tests have only a low association with performance, the employer or educator would be creating inefficiency of production when the tests were used and would be discouraged from using the test. This is simply not true. If individuals really differ in their capacities to perform most tasks, then a test that would rank-order those individuals would be highly valuable. However, most performances in our economy and our education systems are such that they can be performed by large numbers of people. Granted, there are many individual differences; there are also many effective ties. Many people can become good accountants, mechanics, or students. Hence, a blunt or inefficient instrument does not hurt from the point of view of the user, and the economic argument that an inefficient predictor will be discarded is simply not relevant. Any non-perverse predictor will do from the point of view of lowering selection costs and maintaining some kind of adequacy in the pool of applicants accepted.

It's clear that the standardized test itself has often taken the place of a marketplace of people, talents, and employers in allocating

people to social and economic positions. Given that many people can perform many tasks in our economy, and some subset of those people can become well-educated in most things if given the opportunity, we rank-order individuals in terms of test scores, grades, etc., in part for the benefit of low cost usage in selection processes by potential employers or educators. At the same time, we tend to rank-order educational institutions, professions and jobs in terms of their status, the salary they command, and the extent to which they retain for the individual some autonomy over the use of his or her time. We then match and this becomes our surrogate marketplace for labor. Clearly then, employers have reduced their costs, educational institutions have reduced the cost of selecting students and both of them, as well as the professional groups that provide testing tools, have reduced some of the risk that might occur in their respective markets. This is just what public policy is about in the United States as it deals with professions. Costs are to be displaced onto the least organized parties. There is to be a decentralization of policy formation insofar as it affects organized groups. When public authority does exist, the exercise of that authority is delegated to organized professional groups and then used to mitigate further what might be the risk of a real marketplace. Finally, as we have suggested, professional groups defer to each other's judgement.

This deference is becoming the case in court cases involving the use of testing. About the only place in which court cases now occur is when the use of tests appears to be inconsistent with various civil rights laws passed by the Congress and Constitutional guarantees of equal protection. What is interesting is the tendency for courts to increasingly

refer to professional judgements regarding the adequacy of tests. In 1971, in the Griegs vs. Duke Power Company, no mention was made of the APA, AREA, NCME test standards. Great deference was given to the interpretation by public sector agencies of the intent of Congressional legislation. By 1976, in Washington vs. Davis, mention is made of the testing standards and it is suggested that agency interpretation is not to be weighted necessarily more highly than expert judgement. This greater deference to expert opinion occurs as a professional group gets itself together to the point of speaking with one voice; in other words, as it becomes more like a professional guild. Many people in these nascent guilds view this as a saluatory trend. I do not.

Displacement of cost, decentralized usage, delegated public authority and deference to professional groups seems to be inconsistent with the essentially political character of deciding who is to get what through the public sector. When all the preceeding obtains, those well organized groups who are deferred to are among the principal beneficiaries. There is no guarantee that those services or that fairness guaranteed to the public at large in return really occurs. In the case of testing standards, for example, the standards involve the profession's conception of what bias is. However, as we pointed out, that does not exhaust the meaning of bias at all. The external predictive value of a test and the extent of individual inequities that occur through use of the test, principally because of the low cost of its application are not addressed by the professional conception of a bias-free test. Further, the use of even a highly predictive test assumes that what the test predicts is the relevant standard by which to select individuals for educational opportunity or jobs. I think that's unclear.

If we assume that a test of academic performance, say an LSAT, perfectly predicted how well students would do in every given law school, it's still quite unclear that one would want to select those students who scored best. One can think of at least two other criteria of selection. If we are to be economical in our use of our educational facilities, we might want to choose those students who have demonstrated the greatest degree of recent growth, assuming that such a trend would continue. One might then select those students who over the preceeding two years had most increased their test scores, not those who had the highest recent test scores. Alternatively, one could assume that many people, regardless of how well they do in law school, become for the most part, perfectly competant laywers for most types of cases. After all, eventually we're dealing with lawyers and not law school students. We could then assume that we want a modicum of equity with respect to the availability of lawyers to citizens in all economic and social situations. This could be important since we tend to litigate a higher and higher percentage of disputes among individuals and groups in American society. We then go backwards from the distribution of groups and their interests to select those individuals for law school based upon their likelihood of living among and serving particular types of constituencies. So long as they learn the law adequately, this is a perfectly reasonable way to select lawyers if public sector dollars are supplementing the development of lawyers, if public law is being used to delegate authority to a private professional group, and if politically, our representatives decide that legal services are to be provided to all citizens.

Individuals may subject to various of these alternatives for selecting law school students. The resolution of this debate, though, the meeting

of these objections does not occur by using tests. Instead, it occurs in ethical and political discourse and eventually, through appeal to our political system. When we allow the use of standardized tests to implicitly define our criteria of selection for educational opportunity or jobs, we essentially delegate what is a political or ethical decision to a group of individuals who, while not meaning to make that decision, do so as a projection of their own perception of professional norms.

In the preceeding pages, we have mentioned some of the difficulties in the particular use of standardized tests in the development of public policies. Throughout this paper, however, we have attempted to emphasize that what has been happening in the relationship between the use of standardized tests and public policy formation is not unique at all and particularly not unique when professional groups or guilds are involved. The displacement of costs, the disaggregated use of tests, the delegation of public authority, particularly at the state level (and in other countries as well), and the deference given to professional judgements are not unique to the area of testing. Testing is a nascent professional activity when compared with accountancy, law and medicine. The longer standing professional groups have a much greater degree of both formal and detailed structure. In the case of medicine, once doctors were organized at the state level, the only vehicle available to the federal government to pursue public policies regarding the availability of medical care was to spend a lot of money. Since the public sector did not have a lot of control over entry into the medical profession (unlike the U.K.), it could not otherwise influence levels of remuneration or distribution of services. It would be desirable to arrest this development of deference, delegated public authority, and displaced costs

in the case of standardized testing before it goes any further. It would be a shame if we allowed yet another group of professionals to inadvertently specify the sorts of public values that ought to be determined through the political process itself.

REFERENCES

1. Bachrach, Peter, and Baratz, Morton. "Two Faces of Power," APSR, LVI, Dec., 1962, pp. 947-952.

2. Barnes, E.J. "Cultural retardation or shortcomings of assessment techniques" In R. Jones (ed.), Black Psychology. New York: Harper & Row, 1972.

3. Brodinsky, Ben. "Back to Basics, the Movement and its Meaning," in Phi Delta Kappan, March 1977.

4. Burrow, James G. Organized Medicine in the Progressive Era. Baltimore: John Hospkins, 1977.

5. Cardall, C. and Coffman, W.E. A method for comparing the performance of groups on the items in a test (ETS RB 64-61). Princeton, Educational Testing Service, 1964.

6. Cawelti, Gordon. "National Competency Testing: A Bogus Solution," Phi Delta Kappan, Vol 58 (9) May 1978.

7. Cohen, Rosalie A. "Conceptual Styler, Culture, Conflict and Nonverbal Tests of Intelligence," American Anthropologist, 71 (1969), pp. 828-855.

8. Compensatory Education Services, A Report from the National Institute of Education. U.S. Department of Health Education and Welfare. Washington, D.C., July 31, 1977.

9. Congressional Record, October4, 1966-October 6, 1966. pp. 25,027-25,581.

10. Cleary, T.A., and Hilton, T.L. "An Investigation of item bias," Educational and Psychological Measurement, 1968, 28, 61-75.

11. Educational Amendemnts of 1974 (Public Law 93-380).

12. Evaluating Compensatory Education, An Interim Report on the NIE Compensatory Educational Study, The National Institute of Education, U.S. Dept. of Health, Education and Welfare, Washington, D.C., December 30, 1976.

13. Evans, F., and Reilly, R. A study of speediness as a source of test bias (ETS RB 71-131). Princeton: Educational Testing Service, 1971.

14. Ewing, N.J. "Minimum compentency testing and the handicapped: Major issue," The High School Journal, 63, pp. 114-118.

15. Fisher, Thomas H. "Florida's Approach to Competency Testing," Phi Delta Kappan, May 1978.

16. Flynn, J.T., and Anderson B.E. "The effect of test administration procedures on achievement test performance," Journal of Negro Education, 1976, 1, pp. 37-45.

17. Gallup, George H., "Ninth Annual Gallup Poll of the Public's Attitude Toward the Public Schools," September, 1977.

18. Glass, Gene V. "Looking at minimal competency testing"," Education and Urban Society. Vol 12, Vo. I, pp. 47-55.

19. Glass, Gene V. "Minimum Competence and Incompetence in Florida," Phi Delta Kappan. Vol 59, No. 9., May 1978. pp. 602-605.

20. Hall, V.C., and Turner, R.R. "The validity of the 'different language explanation' for poor scholastic performance by black students," Review of Educational Research. 1974, 44, pp. 69-81.

21. "Has Title I Inporved Education for Disadvantaged Students?," National Assessment of Educational Progress, Education Commission of the States, April 1981.

22. Hill, K.T. "The relation of evaluative practices to test anxiety and achievement motivation," Educator, 1977, 19, pp. 15-21.

23. Kelley, E.W. "The Politics of Proficiency..." report to NIE on the origins and consequences of minimum competency exams. 1977.

24. Lazar, Irving et.al. "The Persistance of preschool effects," DHEW, October, 1977.

25. Lazar, Irving et. al. "Lasting Effects after Preschool," HHS, October 1979.

26. Linn, R.L. "Fair test use in selection", Review of Educational Research, 1973, 43, pp. 139-161.

27. Lord, F.M. "A study of item bias, using item characteristic curve theory," Y.H. poortinga (ed.), Basic Problems in Cross-Cultural Psychology. Amsterdam: Swets and Zeitlinger, 1977.

28. Mattleman, Kean. "Minimum Competency, The LEA Perspective," Education and Urban Society, v. 12, November 1979, p. 22.

29. Maw, Carlyle, Item Response Patterns and Group Differences, unpublished Ph.D. Dissertation, University of Chicago, 1979.

30. McConnell, G. Private Power and American Democracy. New York: Knopf, 1966.

31. Merz, W.R. Factor Analysis as a Technique in Analyzing Test Bias. Paper presented at the annual meeting of the California Educational Research Association, Los Angles, 1973.

32. Nairn, A. and associates, The Reign of ETS. Manuscript version of the Nader report on the Educational Testing Service. 1980.

33. Noonan, Robert and Rubin, Larry. "Statewide Minimum Pupil Proficiency Standards" Rutgers University Study done under NIE Grant C-76-10105.

34. Olney, James T. "Minimum Competency Testing in 1978: A Look at State Standards," Legislative Research, 1977, pp. 1-3.

35. Olson, M. The Logic of Collective Action. Boston: Harvard University Press, 1965.

36. Peterson, N.W., and Novick, M.R. "An evaluation of some models for culture-fair selection," Journal of Educational Measurement, 1976, 13, pp. 3-29.

37. Pike, L.W., and Evans, F.R. Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items, CEEB Research Report I. New York: CEEB. 1972.

38. Potthoff, R.F. Statistical Aspects of the Problem of Biases in Psychological Tests. Institute of Statistics Mimeo Series No. 479. Chapel Hill: University of North Carolina, Department of Statistics, 1966.

39. Rudner, L.M. An Evaluation of Select Approaches for Biased Item Identification. Unpublished Doctoral dissertation, Catholic University of America, 1977.

40. Thomas, A., Hetrig, M.E., Dryman, I., and Fernandez, P. "Examiner Effect in IQ Testing of Puerto Rican Working-Class Children," American Journal of Orthopsychiatry, 1971, 41, pp. 809-821.

41. "Three National Assessments of Reading: Changes in Performance, 1970-1980," National Assessment of Educational Progress, Education Commission of the States, April 1981.

42. Tocqueville, A. Democracy in America. Phillips Bradley, ed. Knopf, 1946 edition.

43. Weeres, Joseph G. "School Politics in Thirty Three of the Local Community Areas within the City of Chicago," unpublished Ph.D. dissertation, University of Chicago, 1971.

44. Williams, R.L. "Abuses and Misuses in Testing Black Children," Counseling Psychology, 1971, 2, pp. 62-73.

45. Debra P. v. Turlington. 474 F. Supp. 244 (1979)

46. Griggs v. Duke Power Co., 401 U.S. 424 (1971)

47. Senate Bill 5845, introduced by Sen. Lavalle to the New York State Senate on April 21, 1980.

48. Memorandum of The College Entrance Examiniation Board In Opposition to Senate 5845-B by Mr. Lavalle., by AcWamre, Lochner, Titus, and Williams, P.C. Legislative Council.

49. Memorandum in Opposition to Senate 5845-B. Submitted on behalf of the Associated Medical Schools by Hinman, Straub, Pigors, & Manning, Attorneys at Law, 90 State Street, Albany, NY 12207. on April 25, 1978.

50. Memorandum of Opposition to Senate bill 5845-B. Submitted by the Educational Testing Service, Princeton, NY 08540.

51. Letter from Constance E. Cook, Vice President for Land Grant Affairs, Day Hall, Cornell University to the Honorable Richard A. Brown, Counsel to the Governor, Executive Chamber, Capital Bldg, Albany, New York, 12224 on June 27, 1979, regarding the ten day bill - Standardized Testing (S. 5200-A)

52. Letter from Bruce I. Zimmer, Executive Director, Law School Admission Council 11 Dupont Orcle NW, Washington, D.C. to Sen. Kenneth P. LaValle, Chairman, Committee on Higher Education, NYS Senate, Legislative Office Bldg., Room 805, Albany, NY on February 11, 1980.

53. Testimony of Allan Nairn who works for Ralph Nader before the New York State Senate and Assembly Higher Education Committees on May 9, 1979 regarding Truth-in-Testing.

54. Testimony of Stephen D. Solomon before the New York STate Senate and Assembly Higher Education Committees on May 9, 1979 in Albany, New York.

55. Harrell, William A. The Braziliam Education System: A Summary. Washington: U.S. Government Printing Office, 1970.

56. Havighurst, Robert J., and Gouveia, Aparecida J. Brazilian Secondary Education and Socio-Economic Development. New York: Praeger Publishers, 1969.

57. Huebener, Theodore. The Schools of West Germany. New York: Columbia University Press, 1951.

58. Kneller, George F. The Education of the Mexican Nation. New York: Columbia University Press, 1951.

59. King, Margaret L., and Male, George A. Sweden: Educational Data. Washington: Government Printing Offce, 1965.